DOI：10.13718/j.cnki.xsxb.2022.01.006

基于 L 曲线方法的 Lasso 正则化参数选择①

吴炜明[1,2]，　王延新[1]

1. 宁波工程学院 理学院，浙江 宁波 315211；2. 安徽工业大学 商学院，安徽 马鞍山 243032

摘要：大数据背景下，基于罚函数的正则化方法是高维数据变量选择的重要方法. Lasso 估计是常用的变量选择方法，而 Lasso 正则化参数的取值直接影响选择模型的性能，是正则化方法成败的关键. 针对 Lasso 估计，提出一种新的 L 曲线(LC)准则选择正则化参数. 数值模拟和实际应用表明：相比 CV，GCV，BIC 等准则，LC 准则能够以较高的概率选择真实的模型，并且具有较小的模型误差.

关　键　词：高维数据；变量选择；Lasso；LC 准则；正则化参数

中图分类号：O213　　**文献标志码**：A　　**文章编号**：1000－5471(2022)01－0036－07

Regularization Parameter Selection of Lasso Based on L-curve

WU Weiming[1,2]，　WANG Yanxin[1]

1. School of Science，Ningbo University of Technology，Ningbo Zhejiang 315211，China；

2. Business School，Anhui University of Technology，Ma'anshan Anhui 243032，China

Abstract：In the background of big data，the regularization method based on the penalty function is vital for variables selection of high-dimensional data. Lasso is a common method for variable selection. The value of Lasso regularization parameters directly affects the performance of the selection model，which is the key to the regularization method. Aiming at Lasso，the L-curve criterion for the selection of regularization parameters has been modified，and the new LC criterion been proposed. Through data simulation and practical application，compared with CV，GCV，BIC and other criteria，the LC criterion can select a real model with a higher probability and has a smaller model error.

Key words：high-dimensional data；variable selection；Lasso；LC criterion；regularization parameter selection

大数据时代已经到来，“数据”贯穿了生活的方方面面，在各行各业中都起着举足轻重的作用. 各个领域为了挖掘潜藏的数据价值，对已有数据进行分析建模，但同时也面临着真实场景过于复杂，易出现高维数据的情况. 在变量维数 p 远大于样本量 n 的情况下，传统低维统计分析方法往往显得力不从心. 首先模型的准确性难以得到保证，其次在解释变量大量增加的情况下，模型对于问题的可解释性变差，分析的焦点

① 收稿日期：2020－03－11

基金项目：全国统计科学研究项目(2019LY06)；浙江省自然科学基金资助项目(LY18A010026)；国家级大学生创新创业训练计划项目(201911058025)；宁波市自然科学基金项目(2021J143).

作者简介：吴炜明，硕士研究生，主要从事数据挖掘与分析研究.

通信作者：王延新，副教授，博士.

被模糊，并且在高维变量情况下，模型的复杂度提高，计算量增加，存在一定的求解困难. 因此，在建模过程中，变量选择显得尤为重要.

高维数据变量选择最常用的方法是基于罚函数的正则化方法[1]，它可以同时进行变量选择和参数估计. 稀疏正则化方法的一般框架为

$$\min_{\boldsymbol{\beta}}\left\{l(\boldsymbol{\beta})+n\sum_{j=1}^{p}p_{\lambda}(|\boldsymbol{\beta}_j|)\right\} \tag{1}$$

其中：$l(\beta)$ 为损失函数，$p_{\lambda}(\cdot)$ 为罚函数，λ 为正则化参数. 常用的正则化方法有 Lasso[2]，adaptive Lasso[3]，relaxed Lasso[4]，SCAD[5]，MCP[6] 等. 在实际应用中，上述方法的正则化参数λ 的调节是非常重要的，正则化参数 λ 的选择决定了模型的性能. 目前常采用 CV(交叉验证)[7]，GCV(广义交叉验证)[8]，AIC(赤池信息准则)[9]，BIC(贝叶斯信息准则)[8] 等多种准则选择正则化参数 λ，但是每种方法都有各自的优缺点. CV 方法的预测误差小，但计算量庞大，而且没有完整理论推导，且解释性较差. GCV 方法容易产生过拟合现象[8]，从而不满足变量选择的一致性要求. AIC 准则可以权衡估计模型的复杂度和模型拟合数据的优良性，但也易出现过拟合现象. BIC 准则选择的模型更加接近于真实模型，但是它只考虑了变量选择，参数估计的效果不一定好. Hansen[10] 针对岭回归问题提出最优化参数选择的 L 曲线法. L 曲线方法简单易行，不受模型误差方差的影响，但 L 曲线方法不一定适用于 Lasso 正则化参数的选择.

鉴于以上原因，本文运用 L 曲线的思想，提出一种新的 L 曲线准则(LC) 选择 Lasso 正则化参数. 通过数值模拟，比较 CV，GCV，BIC 与 LC 在 Lasso 方法中模型选择和参数估计的效果. 最后将该方法运用在实际数据中，分析探讨 2019 年 186 个国家经济自由指数的影响因素.

1　Lasso 估计原理与方法

1.1　Lasso 估计

考虑线性模型：

$$\boldsymbol{y}=\boldsymbol{X\beta}+\boldsymbol{\sigma\varepsilon} \tag{2}$$

其中：$\boldsymbol{y}=(y_1,y_2,\cdots,y_n)^{\mathrm{T}}$ 为响应变量；$\boldsymbol{X}=[\boldsymbol{x}_1,\boldsymbol{x}_2,\cdots,\boldsymbol{x}_p]\in\mathbb{R}^{n\times p}$ 为解释变量所组成的样本数据，$\boldsymbol{x}_j=(x_{1j},x_{2j},\cdots,x_{nj})^{\mathrm{T}}$，$j=1,2,\cdots,p$ 为解释变量；$\boldsymbol{\beta}=(\beta_1,\beta_2,\cdots,\beta_p)^{\mathrm{T}}$ 为线性方程的回归系数；$\boldsymbol{\varepsilon}=(\varepsilon_1,\varepsilon_2,\cdots,\varepsilon_n)^{\mathrm{T}}$ 为随机误差，并且 ε_i 服从均值为 0，方差为 1 的独立同分布.

1996 年，文献[2] 提出了 Lasso 方法，通过对回归系数的 L_1 范数进行惩罚来压缩回归系数，并使绝对值较小的回归系数被自动压缩为 0，从而同时实现参数估计和变量选择，基于线性回归的 Lasso 模型为

$$\hat{\boldsymbol{\beta}}=\arg\min_{\beta\in\mathbb{R}^p}\frac{1}{2n}\|\boldsymbol{y}-\boldsymbol{X\beta}\|_2^2+\lambda\cdot\|\boldsymbol{\beta}\|_1 \tag{3}$$

其中：$\boldsymbol{\lambda}\geqslant 0$ 为正则化参数，$\|\cdot\|_2$ 表示 L_2 范数，$\|\beta\|_1=\sum_{p=1}^{p}|\beta_p|$ 为 L_1 范数. Lasso 正则化方法对应的优化问题是凸优化问题，具有稀疏解.

1.2　参数选择方法

正则化参数 λ 的选择决定了模型的性能，因此参数 λ 的选择至关重要. 目前 Lasso 方法常通过 CV，GCV，AIC，BIC 等多种方法来确定参数.

1) CV 方法是一种无假设，可以直接进行参数估计的变量选择的方法. 其思想是在给定样本中，拿出大部分样本进行建模(训练集)，留小部分样本用建立的模型进行预测(测试集)，并计算小部分样本的预测误差，记录误差平方和. 它的优点是预测误差小，但是计算量庞大，而且没有完整的理论依据推导，解释性较差. CV 方法的公式如下：

$$I_{\mathrm{CV}}=\frac{1}{n}\sum_{k=1}^{n}(y_i-\hat{\boldsymbol{\mu}}_{\lambda}^{-k}(\boldsymbol{X}_k))^2 \tag{4}$$

2) GCV 计算过程简单，GCV 具体形式为

$$I_{\mathrm{GCV}}=\frac{\| \boldsymbol{y}-\boldsymbol{X}\hat{\boldsymbol{\beta}}_{\lambda} \|^{2}}{n\left(1-\frac{\hat{f}}{n}\right)^{2}} \tag{5}$$

其中 $\hat{f}$ 是由下式给出的广义自由度:

$$\hat{f}=\mathrm{tr}\{\boldsymbol{X}(\boldsymbol{X}^{\mathrm{T}}+n\sum\lambda)^{\mathrm{T}}\boldsymbol{X}^{\mathrm{T}}\}$$

且 $\sum\boldsymbol{\lambda}=\mathrm{diag}\left\{\frac{\boldsymbol{p}'_{\lambda}(|\hat{\boldsymbol{\beta}}_1|)}{|\hat{\boldsymbol{\beta}}_1|},\cdots,\frac{\boldsymbol{p}'_{\lambda}(|\hat{\boldsymbol{\beta}}_p|)}{|\hat{\boldsymbol{\beta}}_p|}\right\}$. $\sum\lambda$ 的对角元素是罚函数 $p_{\lambda}(\cdot)$ 的局部二次逼近中的二次项系数.

但文献[8]指出GCV方法容易产生过拟合现象,即在参数选择时,λ 容易过小,则非零 $\boldsymbol{\beta}$ 数量就会过多,造成模型的过拟合,从而不满足变量选择的一致性要求.

3) 基于BIC准则的正则化参数选择大致对应于在适当的贝叶斯公式中最大化选择真实模型的后验概率,BIC准则定义如下:

$$I_{\mathrm{BIC}}=\log\left(\frac{\| \boldsymbol{y}-\boldsymbol{X}\hat{\boldsymbol{\beta}}_{\lambda} \|_{2}}{n}\right)+\frac{\log n}{n}\cdot\hat{f} \tag{6}$$

理论上已经证明BIC准则满足模型选择的一致性要求,由BIC准则选择的模型更加接近于真实模型,但是它只考虑了变量选择,参数估计的效果不一定好. 在高维情形下的BIC准则可见文献[10].

2 基于LC准则的正则化参数选择

2.1 岭回归中的L曲线准则

岭回归模型[11]为:

$$\hat{\boldsymbol{\beta}}=\arg\min_{\boldsymbol{\beta}\in\mathbb{R}^{p}}\frac{1}{2n}\| \boldsymbol{y}-\boldsymbol{X}\boldsymbol{\beta} \|_{2}^{2}+\boldsymbol{\lambda}\cdot\| \boldsymbol{\beta} \|_{2}^{2} \tag{7}$$

其中 $\boldsymbol{\lambda}\geqslant 0$ 为正则化参数. 岭估计的罚函数是 L_2 范数,不能把系数压缩到零,因此不能产生稀疏解. 岭参数的选择会在很大程度上影响估计的结果.

文献[12]提出了一种新方法,通过观察由点构成的曲线确定岭回归中的岭参数. 其中横坐标为 $\lambda_i(i=1,\cdots,l$, l 表示预先给定的个数)点处的损失函数 $\| \boldsymbol{y}-\boldsymbol{X}\boldsymbol{\beta} \|_{2}^{2}$ 的对数值,纵坐标表示 λ_i 点处的罚函数值 $\| \boldsymbol{\beta} \|_{2}^{2}$ 的对数值. 通过奇异值分解方法分析了该曲线的所有性质,并指出该曲线上曲率最大的点对应的正则化参数 λ_i 即为最优正则化参数,曲率最大的点记为L-corner. 由于曲线呈现为L形,因此Hansen将这种由残差范数和解范数为坐标点构成的曲线来寻找最优正则化参数的方法称之为L曲线准则. 文献[13]讨论了是否对横纵坐标取对数. 文献[14]介绍了L-corner的数值解法,当选择的正则化方法的正则化参数连续变动时,由残差范数和解范数为坐标构成的L曲线是光滑的,可能是二次可微的,那么L-corner就位于L曲线曲率 $\boldsymbol{\kappa}(\boldsymbol{\lambda})$ 最大处,L-corner处对应的 λ_0 即是最优正则化参数的值. 曲率 $\boldsymbol{\kappa}(\boldsymbol{\lambda})$ 计算公式[14]为:

$$\boldsymbol{\kappa}(\boldsymbol{\lambda})=\frac{\boldsymbol{\rho}'\boldsymbol{\eta}''-\boldsymbol{\rho}''\boldsymbol{\eta}'}{((\boldsymbol{\rho}')^{2}+(\boldsymbol{\rho}')^{2})^{\frac{3}{2}}} \tag{8}$$

其中:$\boldsymbol{\rho}$ 表示残差范数,$\boldsymbol{\eta}$ 表示解范数,$'$ 表示对参数 $\boldsymbol{\lambda}$ 求导.

2.2 Lasso中的L曲线准则

本文试运用L曲线准则在Lasso方法中确定最优正则化参数. 考虑最优化问题(3),构造以($\| \boldsymbol{y}-\boldsymbol{X}\boldsymbol{\beta} \|_{2}^{2}$, $\| \boldsymbol{\beta} \|_{1}$)为坐标点的曲线,其中横坐标为 λ_i 点处的损失函数 $\| \boldsymbol{y}-\boldsymbol{X}\boldsymbol{\beta} \|_{2}^{2}$ 的值,纵坐标表示 λ_i 点处的罚函数值 $\| \boldsymbol{\beta} \|_{1}$ 的值. 找出该曲线的L-corner,该点的正则化参数 λ_i 即为最优参数. 但通过多组不同的数据进行仿真模拟,以($\| \boldsymbol{y}-\boldsymbol{X}\boldsymbol{\beta} \|_{2}^{2}$, $\| \boldsymbol{\beta} \|_{1}$)为坐标点绘制曲线,发现不容易找到曲线拐点(图1). 因此可以认为,对于Lasso正则化方法而言,L曲线准则不容易找出最优的正则化参数 λ_0.

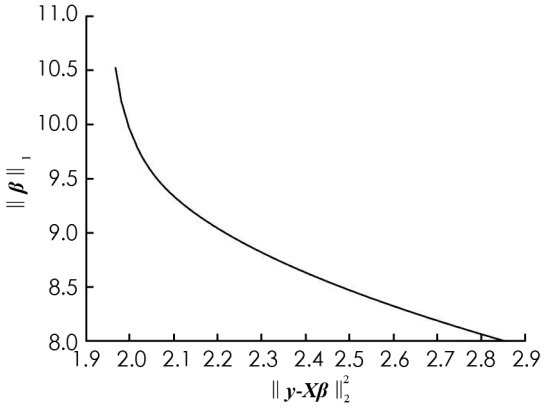

图 1　Lasso 正则化的 L 曲线

鉴于此，本文提出一种适合 Lasso 正则化参数选择的新 L 曲线准则. 所谓 L 曲线准则，是指由坐标为 $(\|\boldsymbol{y}-\boldsymbol{X}\boldsymbol{\beta}\|_2^2, d_0 \cdot \|\boldsymbol{\beta}\|_1)$ 的点构成的光滑曲线，每一个坐标点对应一个正则化参数 λ_i，其中 d_0 表示 Lasso 估计中非零参数的个数. 值得注意的是，以 $(\|\boldsymbol{y}-\boldsymbol{X}\boldsymbol{\beta}\|_2^2, d_0 \cdot \|\boldsymbol{\beta}\|_1)$ 为坐标的点构成的光滑曲线为 L 形状(图 2). 当 $\|\boldsymbol{y}-\boldsymbol{X}\boldsymbol{\beta}\|_2^2$ 与 $d_0 \cdot \|\boldsymbol{\beta}\|_1$ 同时取得最小值(曲率最大) 时的点即为 L-corner，对应的 λ_0 即为最优正则化参数. L-corner 的数值解法与岭回归中 L 曲线准则相同，详情见文献[14]，此处不再阐述.

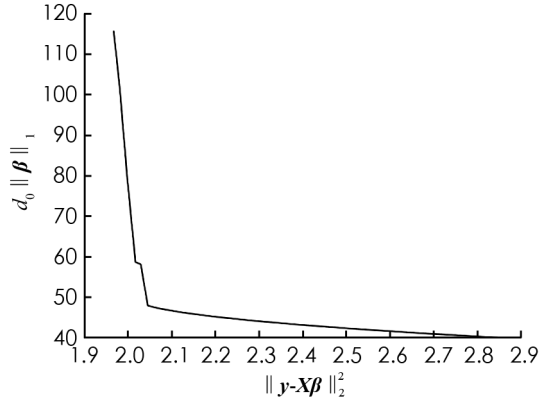

图 2　Lasso 正则化 L 曲线

3　数值模拟与实际应用

3.1　数值模拟

本节通过数值模拟，来比较在 CV，GCV，BIC，LC 下通过 Lasso 正则化方法进行变量选择以及参数估计.

考虑线性模型(2)，取 $\sigma=1,2$，样本与变量数分别取 $n=100,200$，$p=12,20,200,500$ 形成多组随机模拟数据，且 $\boldsymbol{\beta}=(3,1,5,0,0,2,0,0,\cdots)_{1\times p}^{\mathrm{T}}$，$x_i$ 和 x_j 之间的相关系数为 $\mathrm{cor}(j_1,j_2)=0.5^{|j_1-j_2|}$. 在算法上，Lasso 估计采用 ADMM 算法[15]，分别通过 CV，GCV，BIC，LC 选择正则化参数. 重复进行 100 次模拟实验，模拟结果如表 1(低维情形) 和表 2(高维情形) 所示.

为比较估计精确性，需计算模型误差

$$ME(\hat{\boldsymbol{\mu}})=(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta})^{\mathrm{T}}E(\boldsymbol{X}\boldsymbol{X}^{\mathrm{T}})(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}) \tag{9}$$

通过多次的重复试验，用以下指标来评价不同参数选择方法下 Lasso 估计的模型性能. “*MME*”表示模型误差 *ME* 的中位数；“*SD*”表示模型误差 *ME* 的标准差；“*C*”表示 100 次重复实验中非零系数被正确估计为非零个数的均值；“*IC*”表示 100 次重复实验中零系数被错误估计为非零个数的均值；“*Underfit*”表示欠拟合，即在 100 次模拟实验中将非零系数错误估计为零的比例；“*Correctfit*”表示正确拟合，即在 100 次模

拟实验中将非零系数正确估计为非零的比例;"*Overfit*"表示过拟合,即100次模拟实验中选择了所有重要变量并且包含了非零系数的比例.

表1和表2分别展示了低维数据和高维数据两种情况,在不同的随机误差水平下,运用多种变量选择的方法进行Lasso估计.从参数估计误差角度来看,Lasso估计在LC准则下误差比CV方法选择的模型误差小,但是比BIC准则选择的模型误差大,即Lasso估计在LC准则下参数估计的效果介于CV方法和BIC准则之间.从模型的稀疏性角度来看,Lasso估计在LC准则下选择模型较CV,GCV,BIC具有更高的正确拟合比例,具有更低的过拟合比例,即LC准则下的Lasso估计能够选择较稀疏的模型.从变量选择的一致性角度来看,Lasso估计在LC准则下的系数估计效果比CV,GCV,BIC都好,即LC准则下Lasso估计所选择的变量的一致性较好.

表1 低维数据模拟

σ	(n, p)	准则	*MME*	*SD*	*C*	*IC*	*Underfit*	*Correctfit*	*Overfit*
1	$n=100, p=12$	CV	0.173 5	0.121 7	3.00	0.55	0.00	0.59	0.41
		GCV	0.073 1	0.050 8	3.00	3.06	0.00	0.08	0.92
		BIC	0.071 3	0.060 0	3.00	1.06	0.00	0.38	0.62
		LC	0.126 4	0.163 3	3.00	0.13	0.00	0.87	0.13
	$n=100, p=20$	CV	0.198 7	0.110 3	3.00	0.57	0.00	0.65	0.35
		GCV	0.093 6	0.067 5	3.00	4.23	0.00	0.12	0.88
		BIC	0.091 5	0.069 9	3.00	1.15	0.00	0.43	0.57
		MLC	0.151 7	0.169 4	3.00	0.06	0.00	0.94	0.06
	$n=200, p=20$	CV	0.133 8	0.056 9	3.00	0.44	0.00	0.70	0.30
		GCV	0.042 8	0.028 5	3.00	3.25	0.00	0.10	0.90
		BIC	0.045 3	0.036 4	3.00	0.93	0.00	0.43	0.57
		LC	0.081 3	0.088 9	3.00	0.02	0.00	0.98	0.02
2	$n=100, p=12$	CV	0.665 8	0.423 2	3.00	0.45	0.00	0.66	0.34
		GCV	0.272 0	0.218 0	3.00	2.38	0.00	0.18	0.82
		BIC	0.266 6	0.218 0	3.00	0.92	0.00	0.39	0.61
		LC	0.442 8	0.915 2	3.00	0.02	0.00	0.98	0.02
	$n=100, p=20$	CV	0.772 5	0.397 0	3.00	0.64	0.00	0.56	0.44
		GCV	0.308 1	0.254 4	3.00	3.68	0.00	0.11	0.89
		BIC	0.337 0	0.248 4	3.00	1.16	0.00	0.40	0.60
		LC	0.590 2	1.260 8	2.98	0.02	0.02	0.96	0.02
	$n=200, p=20$	CV	0.461 6	0.228 0	3.00	0.36	0.00	0.69	0.31
		GCV	0.172 2	0.129 1	3.00	3.89	0.00	0.15	0.85
		BIC	0.166 4	0.119 4	3.00	0.91	0.00	0.43	0.57
		LC	0.255 3	0.427 4	3.00	0.02	0.00	0.98	0.02

表2 高维数据模拟

σ	(n, p)	准则	*MME*	*SD*	*C*	*IC*	*Underfit*	*Correctfit*	*Overfit*
1	$n=100, p=200$	CV	0.273 5	0.122 8	3.00	1.55	0.00	0.51	0.49
		BIC	0.237 0	0.107 1	3.00	0.44	0.00	0.63	0.37
		LC	0.268 5	0.145 0	3.00	0.09	0.00	0.91	0.09
	$n=200, p=500$	CV	0.185 0	0.072 2	3.00	0.74	0.00	0.69	0.31
		BIC	0.116 5	0.056 9	3.00	0.44	0.00	0.67	0.33
		LC	0.138 7	0.096 3	3.00	0.00	0.00	1.00	0.00
	$n=400, p=1\,000$	CV	0.106 5	0.040 2	3.00	0.48	0.00	0.80	0.20
		BIC	0.063 9	0.028 2	3.00	0.16	0.00	0.85	0.15
		LC	0.071 2	0.048 4	3.00	0.00	0.00	1.00	0.00

续表 2

σ	(n, p)	准则	*MME*	*SD*	*C*	*IC*	*Underfit*	*Correctfit*	*Overfit*
		CV	1.177 2	0.555 4	3.00	1.48	0.00	0.49	0.51
	$n=100, p=200$	BIC	0.894 4	0.544 5	3.00	0.45	0.00	0.64	0.36
		LC	1.109 5	1.311 0	2.98	0.03	0.02	0.95	0.03
		CV	0.665 5	0.309 6	3.00	1.03	0.00	0.67	0.33
2	$n=200, p=500$	BIC	0.473 8	0.252 0	3.00	0.28	0.00	0.78	0.22
		LC	0.565 3	0.405 1	3.00	0.00	0.00	1.00	0.00
		CV	0.431 6	0.166 2	3.00	0.97	0.00	0.80	0.20
	$n=400, p=1\ 000$	BIC	0.268 5	0.123 2	3.00	0.22	0.00	0.81	0.19
		LC	0.313 2	0.141 0	3.00	0.00	0.00	1.00	0.00

3.2　实例分析

本节在 kaggle 平台下载 2019 年世界 186 个国家的经济自由指数的相关数据，该数据集共有 13 个变量，涵盖 186 个国家的 12 项自由指标，从财产权到财务自由，分别为：财产权 X_1；司法效力 X_2；政府诚信 X_3；税收负担 X_4；政府支出 X_5；财政健康 X_6；商业自由 X_7；劳工自由 X_8；货币自由 X_9；贸易自由 X_{10}；投资自由 X_{11}；财务自由 X_{12}；经济自由指数 Y. 对数据进行缺失值和异常值处理，剩下 173 个国家的样本数据. 把经济自由指数作为响应变量，其余 12 个变量作为解释变量，进行实例分析建模.

通过分析，从表 3 可以看出，经济自由指数与其余各因素呈现较强的线性关系，即有线性模型：

$$y_i = \sum_{j=1}^{12} x_{ij}\beta_j + \varepsilon_i,\ i = 0,1,\cdots,173 \tag{10}$$

其中：y_i 表示第 i 个国家的经济自由指数(得分)，x_{ij} 为第 i 个国家的第 j 个变量，ε_i 是均值为 0，方差为 σ^2 的随机误差项.

表 3　线性模型结果

变量	估计	标准差	T 值	$Pr(>\|t\|)$
β_1	8.40E−02	3.26E−04	257.971	<2E−16***
β_2	8.30E−02	3.01E−04	275.615	<2E−16***
β_3	8.34E−02	3.03E−04	275.202	<2E−16***
β_4	8.33E−02	2.29E−04	363.631	<2E−16***
β_5	8.35E−02	1.35E−04	620.092	<2E−16***
β_6	8.33E−02	8.17E−05	1 018.977	<2E−16***
β_7	8.35E−02	2.70E−04	309.680	<2E−16***
β_8	8.33E−02	1.95E−04	427.345	<2E−16***
β_9	8.30E−02	3.03E−04	273.902	<2E−16***
β_{10}	8.31E−02	3.24E−04	256.286	<2E−16***
β_{11}	8.31E−02	1.99E−04	417.092	<2E−16***
β_{12}	8.34E−02	2.35E−04	354.345	<2E−16***
Intercept	7.07E−06	1.18E−05	0.597	0.552

注：*** 表示极其显著.

利用 OLS(最小二乘估计)，CV，GCV，BIC 和 LC 下的 Lasso 估计对该数据进行分析. 变量选择结果如表 4 所示. 从变量选择的数量来看，最小二乘估计 (OLS) 选择了所有的变量，CV 下的 Lasso 罚估计也选择了全部 12 个变量，没有达到变量选择的目的；GCV 和 BIC 准则下的 Lasso 估计分别选择了 11 个和 12 个变量；通过 LC 准则的 Lasso 罚估计选择了 3 个重要变量，分别为 X_3, X_4, X_5，模型也更为稀疏.

4　结论

本文讨论了 Lasso 正则化方法在变量选择和参数估计中的应用，针对 Lasso 正则化提出了 LC 准则，从而更好地确定在不同数据情况下的最优正则化参数. 数据模拟和实际应用的结果都表明，Lasso 估计在

LC准则下能够选择较稀疏的模型，且有较高的概率选择与真实情况相吻合的模型，模型选择效果好. 另外LC准则下的模型的误差较小，参数估计效果好. 本文的LC准则同样可以推广到非线性模型中.

表4 不同方法下的参数估计结果

变量	OLS	CV	GCV	BIC	LC
β_1	13.136 9	4.942 5	12.983 0	11.639 1	0.000 0
β_2	3.824 0	4.162 5	3.815 1	3.588 0	0.000 0
β_3	1.896 6	3.437 4	2.031 5	2.550 3	5.873 1
β_4	−5.336 1	5.501 7	−5.202 7	−5.166 0	−2.727 3
β_5	1.279 9	5.102 2	1.019 8	0.000 0	−1.497 6
β_6	6.693 7	5.972 8	6.661 5	6.416 5	0.000 0
β_7	0.469 4	5.418 7	0.000 0	0.000 0	0.000 0
β_8	3.843 6	4.987 8	3.617 7	2.220 6	0.000 0
β_9	−19.985 1	5.823 0	−19.511 3	−17.388 7	0.000 0
β_{10}	−16.914 9	5.876 9	−16.314 0	−13.797 4	0.000 0
β_{11}	6.178 7	4.865 3	5.983 4	5.023 8	0.000 0
β_{12}	7.716 8	4.523 5	7.696 2	7.562 6	0.000 0

参考文献：

[1] 曾津，周建军. 高维数据变量选择方法综述 [J]. 数理统计与管理，2017，36(4)：678-692.

[2] TIBSHIRANI R. Regression Shrinkage and Selection via the Lasso [J]. Journal of the Royal Statistical Society：Series B (Methodological)，1996，58(1)：267-288.

[3] ZOU H. The Adaptive Lasso and Its Oracle Properties [J]. Journal of the American Statistical Association，2006，101(476)：1418-1429.

[4] MEINSHAUSEN N. Relaxed Lasso [J]. Computational Statistics & Data Analysis，2007，52(1)：374-393.

[5] FAN J Q，LI R Z. Variable Selection via Nonconcave Penalized Likelihood and Its Oracle Properties [J]. Journal of the American Statistical Association，2001，96(456)：1348-1360.

[6] ZHANG C H. Nearly Unbiased Variable Selection under Minimax Concave Penalty [J]. The Annals of Statistics，2010，38(2)：894-942.

[7] ALLEN D M. The Relationship between Variable Selection and Data Agumentation and a Method for Prediction [J]. Technometrics，1974，16(1)：125-127.

[8] WANG H，LI R，TSAI C L. Tuning Parameter Selectors for the Smoothly Clipped Absolute Deviation Method [J]. Biometrika，2007，94(3)：553-568.

[9] ZOU H，HASTIE T，TIBSHIRANI R. On the “Degrees of Freedom” of the Lasso [J]. The Annals of Statistics，2007，35(5)：2173-2192.

[10] CHEN J，CHEN Z. Extended Bayesian Information Criteria for Model Selection with Large Model Spaces [J]. Biometrika，2008，95(3)：759-771.

[11] HOERL A E，KENNARD R W. Ridge Regression：Biased Estimation for Nonorthogonal Problems [J]. Technometrics，1970，12(1)：55-67.

[12] HANSEN P C. Analysis of Discrete Ill-Posed Problems by Means of the L-Curve [J]. SIAM Review，1992，34(4)：561-580.

[13] HANKE M. Conjugate Gradient Type Methods [M]//Conjugate Gradient Type Methods for Ill-Posed Problems. Englewood：Chapman and Hall/CRC，2017：7-34.

[14] HANSEN P C，O’LEARY D P. The Use of the L-Curve in the Regularization of Discrete Ill-Posed Problems [J]. SIAM Journal on Scientific Computing，1993，14(6)：1487-1503.

[15] ZHU Y Z. An Augmented ADMM Algorithm with Application to the Generalized Lasso Problem [J]. Journal of Computational and Graphical Statistics，2017，26(1)：195-204.

责任编辑 张枸